**UNITED STATES DISTRICT COURT**
**DISTRICT OF MASSACHUSETTS**

| | |
|---|---|
| SCANSOFT, INC.,<br><br>Plaintiff<br><br>v.<br><br>VOICE SIGNAL TECHNOLOGIES, INC., LAURENCE S. GILLICK, ROBERT S. ROTH, JONATHAN P. YAMRON, and MANFRED G. GRABHERR,<br><br>Defendants | C.A. No. 04-10353-PBS |

# SCANSOFT'S *MARKMAN* BRIEF ON CLAIM CONSTRUCTION OF U.S. PATENT 6,594,630

Lee Carl Bromberg, BBO # 058480
Robert M. Asher, BBO # 022865
Erik Paul Belt, BBO # 558620
Lisa M. Fleming, BBO # 546148
Jack C. Schecter, BBO # 652349
Rebecca Hanovice, BBO # 660366
BROMBERG & SUNSTEIN LLP
125 Summer Street
Boston, MA 02110-1618
Tel: (617) 443-9292

*Counsel for ScanSoft, Inc.*

Dated: June 3, 2005

# TABLE OF CONTENTS

# TABLE OF AUTHORITIES

## CASES



## STATUTES & RULES

## INTRODUCTION

United States Patent No. 6,594,630 ("the '630 patent") is entitled "Voice-Activated Control for Electrical Device" and, as the title suggests, is directed to apparatuses and methods that use voice commands to control power to appliances such as lighting fixtures. **Exh. 1**, '630 patent at col. 1, *ll.* 5-10. For instance, the patent specification gives the example of turning on a lamp simply by saying "LIGHTS<pause>ON." **Exh. 1**, '630 patent at col. 3, *ll.* 54-55.

Controlling appliances via voice commands, however, had long been known before the application for the '630 patent was filed in 1999. The PTO examiner who issued the '630 patent stated as much. *See* **Exh. 2**, Prosecution History of '630 Patent at PH68 ("controlling electrical devices with voice has been long practiced . . ."). So while the '630 patent concerns the use of voice commands to turn on lamps or other appliances, the purported novelty lies elsewhere.

As the speech recognition expert, Bruce Balentine, discusses in his tutorial on voice activated dialing, one of the major concerns in the speech recognition industry is reducing errors in speech recognition--*e.g.*, mistaking the command "Call Home" for "Call Office." *See* Balentine I at ¶¶ 18, 21, 49-53.[1] A specific technique of error reduction is a prominent feature of the '630 patent and provides its claimed patentability. Specifically, the '630 patent explains that "wordspotting" is the process of recognizing a command (such as a "keyword") embedded in unknown speech (*i.e.*, words not in the programmed vocabulary of the system) or that is spoken over background noise (*e.g.*, conversations, sirens, etc). *See* **Exh. 1**, '630 patent at Col. 1 at *ll.*

---

[1] Mr. Balentine previously submitted a declaration in support of ScanSoft's *Markman* brief on the '966 patent. Mr. Balentine is now submitting a second declaration to discuss VST's '630 patent. For clarity, Mr. Balentine's first declaration, dated May 6, 2005, is cited as "Balentine I," while the second declaration, concerning the '630 patent, is cited as "Balentine II."

39-46. The patent calls errors in wordspotting "false alarms." Col. 1 at *ll.* 57-60. One object of the '630 patent is to improve wordspotting and thus reduce false alarms. Col. 3, *ll.* 50 *et seq.*

One known way to improve wordspotting, according to the patent, is to increase the number of syllables in the command. By lengthening the command (*e.g.*, by requiring more words, such as "PLEASE TURN THE LIGHTS ON," rather than simply "LIGHTS ON"), the speech recognition system can better decipher the utterance "because more information is available for decision making." Col. 1 at *ll.* 50-54. The problem with adding words, however, is that "it is more convenient for users to memorize and say short commands." *Id.* at *ll.* 56-57.

The '630 patent attempts to solve this problem by making the commands longer without adding words. To do so, the patent requires the insertion of a pause between the command keywords. According to the patent, the speech recognition algorithm treats the pause as part of the command phrase. By inserting a pause, the two-syllable command "LIGHTS ON" becomes a three-syllable command, akin to "TURN LIGHTS ON." The patent alleges that this method improves wordspotting without having the user memorize more words. *See* Col. 3 at *ll.* 50-60.

Understanding why this pause is necessary and how the speech recognition system detects the pause (and thus better spots command words) is helpful to construing the disputed terms "pause portion," "syllable," "spectral content," and "dynamic." In addition, the parties also dispute the term "activating." The parties dispute other terms, but construction of the above terms may alleviate the need to construe those others.

## THE DISPUTED CLAIMS

VST asserts Claims 7 and 16 of the '630 patent. Both claims are independent method claims and share many of the same disputed claim terms. Claim 7 is directed to a method for activating an electrical device through a voice command. Claim 7 reads as follows:

7. A method of activating an electrical device through at least one audio command from a user, the method comprising the steps of:

recording speech recognition data having a command word portion and a pause portion, each of the speech recognition data portions being at least one syllable in length;

receiving at least one audio command from a user, the at least one audio command having a command word portion and a pause portion, each of the audio command portions being at least one syllable in length;

comparing said command word portion and said pause portion of said at least one received audio command with said command word portion and said pause portion, respectively, of said speech recognition data;

generating at least one control signal based on said comparison;

controlling power delivered to an electrical device in response to said at least one control signal for operating the electrical device in response to said at least one received audio command;

analyzing the pause portion of the received audio command for spectral content; and

preventing operation of the electrical device when the spectral content is dynamic.

The method of Claim 7 works as follows. First, a command (including both word and pause portions) is recorded. This is the "training" process that Mr. Balentine discusses in his tutorial. *See* Balentine I at ¶ 34. The system stores this sample "speech recognition data," which becomes the model for comparison with audio commands spoken by the users.

After the system is trained (thus establishing the "speech recognition data"), the system is ready to receive commands. The command must have a "command word portion" and a "pause portion." In the example used throughout the patent, the command word portion is "LIGHTS ON," while the pause portion is the momentary stop (*e.g.*, silence) that the speaker inserts between "LIGHTS" and "ON." The claim requires that each portion of the command--both words and pause--must be "at least one syllable in length." The parties dispute the meaning of "pause" and "syllable."

The speech recognizer then compares the word and pause portions of the command with the word and pause portions previously stored as speech recognition data. This step is basic to any speech recognition system, as Balentine explained. *See* Balentine I at ¶¶ 33-42. The patent likewise illustrates the training and comparison steps. *See* Col. 8. *ll.* 34-43 ("Speech recognition is generally performed by comparing the features of an unknown utterance with the features of known words. . . . ."). Based on this comparison (*i.e.,* the recognition of a valid command), the system generates a signal. The signal, in turn, tells the system to turn on power to the appliance.

But there are two other steps that must occur before the system can turn on the lamp. The speech recognizer is constantly listening for commands and attempting to distinguish them from background noise and conversation. Thus, the recognizer must first verify that the speaker has paused between command words to confirm that a command has, indeed, been spoken. The recognizer analyzes the pause portion for "spectral content" (*i.e.*, the collection of frequencies that comprise sound) and prevents operation of the lamp if the spectral content is "dynamic" (*i.e.*, changing). *See* Balentine II at ¶¶ 12-15. Dynamic spectral content indicates that there are sounds where the pause should be and, therefore, a command was not uttered. *See* Col. 10, *ll.* 43-55. Stable spectral content indicates relative silence--*i.e.*, a pause. *See* Balentine II at ¶ 15.

Claim 16 is similar to Claim 7, except that instead of requiring a user to speak only a command and pause portion, it instead requires "first and second command word portions and a first, second and third pause portions" of the command. In other words, the user inserts three pauses--one before the first word, one after the second word, and one in between the words. The patent gives the example of "<pause> LIGHTS <pause> ON <pause>." Col. 10, *ll.* 10-17. In addition, this claim does not include the "analyzing" and "preventing" steps of Claim 7. As argued in Section IV below, this claim is fatally indefinite and thus cannot be fully construed.

## ARGUMENT

### I. THE '630 PATENT DEFINES THE "PAUSE PORTION" AS A "SYLLABLE" OF AT LEAST ABOUT 200 MILLISECONDS

In most instances, claim terms receive their ordinary meaning to those of ordinary skill in the art and are not limited to preferred embodiments in the specification. But the Federal Circuit also holds that intrinsic evidence (the specification and prosecution history) can narrow or qualify a particular claim term in certain instances, such as when the intrinsic evidence clearly and manifestly defines a term, disclaims scope, or provides needed clarity to otherwise vague terms. *See e.g., Alloc, Inc. v. Int'l Trade Comm'n,* 342 F.3d 1361, 1368-69 (Fed. Cir. 2003).

In particular, the intrinsic evidence "provides the technological and temporal context to enable the court to ascertain the meaning of the claim to one of ordinary skill in the art at the time of invention." *V-Formation, Inc. v. Benetton Group SpA*, 401 F.3d 1307, 1310 (Fed.Cir. 2005); s*ee also Vitronics Corp. v. Conceptronic, Inc.,* 90 F.3d 1576, 1582 (Fed. Cir. 1996) ("The specification acts as a dictionary when it expressly defines terms used in the claims or when it defines terms by implication"). In this case, the inventors of the '630 patent have expressly and manifestly defined the terms "pause" and "syllable." Indeed, these terms would lack sufficient clarity and definiteness without the definitions in the specification.

#### A. The Claim Requires a "Pause Portion" to Be "at Least One Syllable in Length"

Both Claims 7 and 16 require that the "pause portion" be "at least one syllable in length." The parties dispute the meaning of "pause portion" and "syllable." A "pause portion" is a momentary stop in speech. As seen in Subsection B below, the patent defines a pause in this fashion and, indeed, the claimed method would not work otherwise. *See* Balentine II at ¶ 11. VST, however, asserts that a pause portion is merely that portion of the command that is not a

word portion. *See* VST's *Markman* brief at 10. But it is hardly a definition to say, simply, that one thing is what the other is not.

As argued in Subsection C, the '630 patent defines "syllable" to have a duration of at least 200 milliseconds ("msec"). That definition is in keeping with the underlying basis of all speech recognition systems--that speech must be parsed into units of time. *See* Balentine I at ¶¶ 25, 29-31. VST, however, asserts a circular definition, which is that a syllable is "within the duration range for one syllable of speech." VST's *Markman* brief at 9. That definition would leave the claim indefinite. *See* Balentine II at ¶¶ 9-10 ("Given the widely variable lengths of syllables, from a practical perspective, . . . claims 7 and 16 of the '630 patent require a syllable-length speech element having at least a duration set above some minimum threshold").

The claim wording suggests that a syllable must have some predefined, minimum length. The claims do not read that a pause may be any random length. Rather, the claims read that a pause must be "at least one syllable in length." The specification and prosecution history show that the inventors defined pauses to be artificial, deliberate stops lasting for at least 200 msec.

**B. A Pause Is an Intentional, Momentary Stop in Speech**

**1. The Specification Defines the Pause**

The '630 patent defines pauses as "unnatural breaks in sound." Col. 10, *ll.* 60-63.[2] That definition complies with the stated technique of the claimed invention, which is the intentional, deliberate, artificial insertion of pauses into speech to improve wordspotting accuracy by adding syllables to commands without adding words. VST argues that a "pause is simply the natural

[2] Specifically, the patent states that "[t]he unknown utterances [spoken by a user] must not only have the sound sequences correct, but must also have the unnatural breaks in sound (the pauses) at the correct time in order to create a competitive score for triggering the system." Col. 10, *ll.* 60-63 (emphasis added). Use of the word "must" demonstrates that this definition is not just a preferred embodiment but rather applies to all embodiments.

space between command words." VST's brief at 10. But the specification belies VST's contention. The specification shows that natural spaces between words will not meet the invention--the pauses must be deliberate and of predefined length for the system to work.

For this reason, the patent insists that "It is not only necessary to separate the voice commands from the background noise in order to issue a control command . . .." Col. 8, *ll.* 22-27 (emphasis added). That is where the pauses come into play. They help separate voice commands from the background noise. *See* Balentine II at ¶ 5 (". . . the longer a command phrase is, the more distinct that phrase will be from normal conversation or background noise and the more accurate the speech recognizer will be in spotting it.").

The deliberate pause also furthers another function of the claimed invention, which is to distinguish actual commands from normal conversation. For example, if a user says to a friend, "the lights are on," the speech recognizer could interpret that as the command "LIGHTS ON." By deliberately inserting a pause of at least a minimum length, however, the user ensures that the command will be recognized. *See* Balentine II at ¶¶ 4-5 ("By requiring a user to include distinct pauses between command words, the speech recognizer of the '630 patent lengthens the command, resulting in greater accuracy."). Conversely, the recognizer will not mistake command words coincidentally embedded in speech with actual commands. *Id.*

In distinguishing the wordspotting methods of the prior art, the patent notes that the "intentional use of speech pauses in wordspotting is reminiscent of the early days of automatic speech recognition . . . where algorithmic limitations required the user to briefly pause between words." Col. 1, *l.* 66 - Col. 2, *l.* 5 (emphasis added). In other words, the patent is conceding that the intentional pauses used in the claimed invention are like (*i.e.*, are "reminiscent" of) the "discrete dictation"/"isolated word" recognition systems of the prior art.

After conceding that the use of pauses is reminiscent of the prior art, the patent attempts to distinguish the prior art pauses. According to the patent, the difference is not in the length of the pause, or whether its insertion is intentional, but rather how the speech recognizer treats those pauses: "In this prior art technique, the pause is not analyzed and therefore is not used in the pattern classification process." Col. 2, *ll.* 17-19. In other words, the difference is not how long the user pauses (as VST appears to suggest) but rather how the speech recognizer interprets the pauses. According to the patent, the prior art "discrete dictation" systems used pauses only to detect the endpoints of words and then discarded or ignored the pauses when interpreting the commands. In contrast, the '630 patent asserts that the recognizer treats the pauses as part of the audio command itself. The speech recognition system does not discard the pauses but rather uses them to make the keywords longer for improved wordspotting. The '630 patent makes this alleged difference from the prior art explicit:

> Accordingly, it is important to note the following key differences between the use of speech pauses in the present invention and in the prior art isolated-word recognition:
>
> > 1) The invention treats the speech pauses as part of the keywords and as such treats them just like any other speech sound. Thus, the particular spectral qualities of the input signal during speech pauses are essential for the keyword to be correctly detected. In contrast, the prior art isolated-word recognition discards speech pauses during a pre-processing step; and
> >
> > 2) The purpose of the speech pauses in the present invention is to make the keywords longer rather than to simplify endpoint detection. In fact, no explicit end-point detection is performed at all in the present invention.

Col. 3, *l.* 62 - Col. 4, *l.* 10.

Thus, the specification contradicts VST's argument that the pauses used in the methods of the '630 patent are somehow different from pauses used in the prior art. The speech recognition algorithms of the prior art and of the '630 patent use the pauses differently. But to

the user, the pauses of the prior art discrete dictation systems are the same as those used in the '630 patent--they are artificial, intentional, deliberate, unnatural breaks in sound.

Contrary to VST's argument, the '630 patent does indeed require the user "artificially to stop speaking in between each word." VST's brief at 10. If the user did not artificially stop speaking for a minimum span of time, the speech recognizer would not be able to distinguish normal conversation coincidentally using the command words (*e.g.*, "Hey, Bill, are the lights on in the other room?") from actual command words intended to turn on the lights (*e.g.*, "LIGHTS <pause> ON"). That is, if normal breaks between words were always treated as pauses, then the method would not work. *See* Balentine II at ¶ 11 ("As the duration of the syllable-length pause decreases, the additional measure of accuracy gained by including that pause portion in the command phrase also decreases.").

As argued in Subsection C, to ensure that the pauses sufficiently lengthen the keywords, and to ensure that the speech recognizer correctly distinguishes commands from other sounds, the patent imposes a minimum duration of 200 msec.. *See* Balentine II at ¶ 10. Thus, the user pauses just has he or she would have in the early days of discrete dictation systems.

**2. The Prosecution History Supports ScanSoft's Definition**

The prosecution history also shows that the inventors did not distinguish the prior art pauses based on their length or artificial nature. In rejecting the claims based on the prior art, the PTO examiner noted that the "characteristic of speech consisting of words separated by pauses is notoriously well known in the art of speech signal processing and would have been obvious to a person of ordinary skill in the art at the time of the invention." **Exh. 2,** Prosecution History at PH 68. The examiner also noted that the *Bellergarda et al.* reference describes the use of

command words separated by pauses. *Id.* The *Bellergarda* prior art patent describes pauses as having a minimum duration of 300 milliseconds:

> According to one embodiment of the present invention, different commands are separated from one another based on detected pauses between words. In this embodiment, if silence is detected for greater than a threshold period of time, then the silence is interpreted as a pause between commands. Typical values for this threshold period of time range from 300 milliseconds to 3 seconds.

**Exh. 3,** U.S. Pat. No. 6,208,971 to Bellergarda *et al.* at Col. 3, *ll.* 24-30 (emphasis added).[3]

The inventors of the '630 patent, however, did not distinguish *Bellergarda et al.* based on the length of the pause. Indeed, the inventors did not distinguish *Bellergarda* at all--in effect conceding that the pauses are the same as those in the prior art. Rather, the inventors considered the PTO examiner's obviousness rejections to be "moot in view of the amendment of the claims presented in the Response." **Exh. 2,** Prosecution History at PH 91. The subsequent claim amendments did not change the definition of "pause" or "syllable." Rather, the amendments concerned another limitation of the present claims--namely, the "analyzing" and "preventing" steps now present in Claim 7. *See id.* at PH 93 and 96 (underlined material indicating the amendments). The PTO examiner stated that the "preventing" step distinguished the prior art, not the use of syllable-length pauses. *Id.* at PH 73 (detailing "allowable subject matter").

### C. A Syllable is At Least 200 Milliseconds

#### 1. The Specification Expressly Defines "Syllable"

As demonstrated above, for the claimed method to work properly, the pause must be a deliberate, unnatural break in sound that is different from a normal break between words in everyday conversation. *See* Balentine II at ¶ 11. Indeed, the patent specifies that a pause model

[3] The Federal Circuit has affirmed that prior art cited in a patent, like *Bellergarda et al.*, is treated as intrinsic evidence. *See V-Formation*, 401 F.3d at 1311.

(the pause portion used to train the system) must be "of predefined duration between keywords in the voice commands." (Both claims require that the pause model in the speech recognition data, like the pause portion of the audio command, be "at least one syllable in length.") Of course, the patent recognizes that the tempo of speech may differ from speaker to speaker (giving the example of "run" versus "ruuuuuuun" at Col. 3, *ll.* 9-11) and that, therefore, the length of the pause may also vary. But to ensure proper operation, the patent specifies a minimum duration for the pause:

> The duration of the pause model **164** between each command word may vary depending on the particular speaking style of the user(s), but should be at least one syllable (about 200 msec.) in length.

Col. 10, *ll.* 5-8 (emphasis added).

The patent then explains how the "pause model" recorded by the user should be configured in the preferred Hidden Markov Model recognition algorithms:

> In order to impose a minimum pause duration of about 200 msec, the pause model **164** needs to contain at least N silence states (represented by $s_i$ in FIG. **4**) where
>
> $N$=minimum pause duration/frame update=200 msec/10 msec=20.

Col. 10, *ll.* 22-26 (emphasis added).

The use in these passages of the words "should be," "needs to," and "impose" distinguish these required embodiments from mere preferred embodiments. In contrast, the inventors consistently use words like "may," "can," and "preferably" when they desire to differentiate permissive from required embodiments *See, e.g.*, Col. 6, *ll.* 35-40 ("The micro-controller **14** preferably includes an 8-bit or 16-bit MCU . . ."); Col. 6, *ll.* 66-67 ("The audio input **12** may be in the form of a microphone or similar device . . ."); Col. 8, *ll.* 20-22 ("The sound data may contain a combination of voice commands . . ."); Col. 10, *ll.* 43-47 ("The presence of a pause is preferably determined by . . ."); Col. 13, *ll.* 19-30 ("The threshold value S

can be either a discrete . . . Although a trim potentiometer is preferred, the threshold adjusting means can take the form of . . ."). But in the above-quoted passages, the inventors declined to use such permissive words.

This use of prescriptive words like "must" or "should be" or "needs to"--as opposed to permissive words like "may" or "can" or "preferably"--shows that the inventors have expressly and manifestly defined or narrowed the term "syllable" to at least 200 msec. *See, e.g., Alza Corp. v. Mylan Labs. Inc.*, 391 F.3d 1365, 1371 (Fed. Cir. 2004) (limiting the term "skin permeable form" of a particular drug compound used in a transdermal patch to the base form of the compound, rather than its acidic form, because the specification stated, *inter alia*, "the drug should be incorporated in the transdermal therapeutic system in the form of the base") (emphasis added); *cf. Prima Texk II, L.L.C. v. Polypap, S.A.R.*L., 318 F.3d 1143, 1150 (Fed. Cir. 2003) (declining to limit claim to preferred embodiment because "[t]he written description only states that the floral holding material 'may be' (not must be) the type of material commonly referred to in the art as floral foam or soil").

Requiring the pause to be at least a minimum length also makes sense so that the system developers know how to configure the recognition algorithms to recognize pauses. The above-quoted passage shows that a pause model must be translated into a certain number of frames (in this case 20 frames, each frame representing about 10 msec.). As Mr. Balentine discusses in his tutorial on speech recognition, the "front end" processor converts each second of sound into 100 frames, each representing 10 msec. of the sound slice. Each frame is a snapshot of the spectral quality of sound at any point in time. Balentine I at ¶¶ 29-30. The recognition algorithms, such as Hidden Markov Models or Template Matching systems, can then use these frames to interpret the sounds. *Id.* at ¶¶ 33-42. But this recognition depends on the fact that the underlying basis for

speech recognition is time. That is, sounds are divided into slices of time, and because time only moves forward, the speech can be reconstructed faithfully. *Id.* at ¶ 25. Thus, one of ordinary skill who wants to build a device to turn on lamps according to the '630 patent must base the pause models on a definite unit of time, a definite number of frames. *See* Balentine II at ¶ 10. A "syllable," standing alone, without further quantification, does not provide the necessary definiteness for the recognition algorithms. *See* Balentine II at ¶¶ 8-9.

Of course, a pause may be longer than one syllable (*e.g.*, to account for a speaker with a slower or more deliberate speaking style), and thus Claims 7 and 16 require that the pause be "at least" a syllable. But the pause cannot be less than one syllable, and, as defined in the '630 patent, a syllable must have a minimum duration of at least about 200 milliseconds.[4]

**2. The Prosecution History Confirms this Definition**

The prosecution history also shows that a "syllable" must be 200 msec. or longer. To overcome rejection of the claims, the inventors pointed to the definition in the specification that a syllable be at least about 200 msec. As such, the inventors expressly disclaimed a broader definition of syllable. *See, e.g., C.R. Bard, Inc. v. U.S. Surgical Corp.*, 388 F.3d 858, 868 (Fed. Cir. 2004) (construing patent to require that surface of surgical device had to be pleated because, during reexamination, in response to the examiner's rejection, the applicant pointed to the passage in the specification disclosing a device that is pleated).

Specifically, in an office action dated December 3, 2001, the PTO examiner rejected the original claims because, he asserted, defining the pause as "at least one syllable in length" was

[4] VST points to its expert's statement that "it is not possible to restrict the term 'at least one syllable' to a specific measure in time." Wooters Declaration at ¶ 4. But that testimony contradicts the intrinsic evidence and should be disregarded. *See C.R. Bard, Inc.*, 388 F.3d at 862 ("our caselaw suggests that extrinsic evidence cannot alter any claim meaning discernible from intrinsic evidence").

indefinite and that, therefore, a "more appropriate unit to designate measurement and duration" was needed. **Exh. 2,** Prosecution History at PH65. In response, the inventors argued that a syllable "is a clear alternative for using the units of time (such as minutes) for the same purpose." The inventors then directed the PTO examiner to their discussion in the specification in which a syllable is defined as a unit of time:

> This technique has been fully explained and further developed by the present application. In this respect, please refer, for example, to the paragraph bridging pages 19 and 20 of the originally filed application. For example, it is specified their that "The duration of the pause model **164** between each command word may vary depending on the particular speaking style of the user(s), but should be at least one syllable (about 200 msec.) in length."

**Exh. 2,** Prosecution History at PH 88-89.

In other words, the inventors argued that "syllable," which the PTO rejected as not describing a duration of time, expressly connotes a unit of time (about 200 msec.). By linking this description to the use of "syllable" in the claims, the inventors made clear that 200 msec. is not merely a preferred embodiment but rather is a required part of the invention as claimed.

## II. THE METHOD OF CLAIM 7 VERIFIES PAUSES BY DETECTING WHETHER THE "SPECTRAL CONTENT" IS CHANGING ("DYNAMIC")

Claim 7 requires the step of "preventing operation of the electrical device when the spectral content [of the pause portion] is dynamic." The parties essentially agree on the meaning of the individual terms "spectral content" and "dynamic" but differ on the combination of these words in the claim. ScanSoft contends that this limitation means what it says--preventing operation when the "spectral content" of the pause portion of the command is changing. That interpretation makes sense because, as shown below, silence (*i.e.*, a pause) is characterized by a relative lack of spectral activity (*i.e.*, stable spectral content), while noise or speech is characterized by dynamic spectral content--*i.e.*, changes in the frequencies that comprise sound.

The purpose of this step is to prevent the speech recognition system from mistaking, *e.g.*, normal conversation or background noise for a true command. The system should activate the appliance only when a pause is detected, showing that a command has actually been spoken.

VST, however, adds a new concept not stated in the claim: preventing operation when the "spectral content is changing more than background noise would be expected to change." VST's Markman brief at 14. In other words, VST contends that the claim requires the speech recognition software to compare the spectral content of the pause with the spectral content of background noise to determine whether a pause has occurred. The claim, however, says nothing about comparing the spectral content of the pause to "background noise" or anything else.

The parties essentially agree that "spectral content" refers to the collection of different frequencies that constitute a particular sound at a particular time. For example, ScanSoft's expert, Mr. Balentine, previously explained in his tutorial that the "front end" signal processor divides sound into frames, each frame representing a slide of the sound spectrum, which is analogous to a prism breaking down light into colors of the light spectrum. *See* Balentine I at ¶¶ 29-31. Based on this prior discussion, Mr. Balentine describes "spectral content" as analogous to "color" in light. *See* Balentine II at ¶ 13. As Mr. Balentine explains, sounds are vibrations in the air around us. When objects vibrate, they vibrate at multiple frequencies, generating harmonics and overtones producing a sound that is actually made up of a concurrent mixture of many different sounds at different frequencies. *Id.* The sum of all of these sounds is perceived by a human ear as the "timbre" or the "color" of the sound and is measured over time as the spectral content of the sound. *See id.* at ¶¶ 13-14. VST's expert, Wooters, describes "spectral content" in similar terms. *See* Wooters Declaration at ¶ 5 ("the amount of energy at each frequency over a specified period of time"). The parties also agree that the term "dynamic" means, simply,

changing. *See* Balentine II at ¶ 18 ("a spectral feature that varies over time is referred to as a 'dynamic' spectral feature"); Wooters Declaration at ¶ 6 ("'Dynamic' generally means changing"); VST's *Markman* brief at 14 ("Both parties agree that 'dynamic' means changing").

The concept of comparing portions of the command to background noise, however, does not appear in Claim 7 but does appear in other claims. For example, dependent Claim 9 includes the step of "comparing the background noise data to the at least one received audio command." Claim 10 includes the steps of "ascertaining the energy content" of the audio command and the background noise and then "comparing the first and second energy contents and generating an energy comparison value." Claim 3 likewise requires "means for comparing the energy of the command word portion to the energy content of the background noise data . . .." Under the doctrine of claim differentiation, these express "comparing" steps found in dependent claims cannot be read into Claim 7, as VST attempts to do. *Liebel-Flarsheim Co. v. Medrad, Inc.*, 358 F.3d 898, 910 (Fed. Cir. 2004).

The specification and prosecution history also support this construction. For example, the specification explains that a silence is a "stationary condition." Col. 9 at *ll.* 4-8. Mr. Balentine agrees. He explains that, within the context of the claims of the '630 patent, relatively stable spectral content indicates a pause (*i.e.*, relative silence), while dynamic spectral content indicates sounds. *See* Balentine II at ¶¶ 16-18.

The patent specification also describes the use of dynamic spectral activity without any reference to background noise:

> The presence of a pause is preferably determined by analyzing both the spectral content and the energy content of the pause before and/or after the detection of a keyword, depending on the particular sequence of pauses and keywords. If dynamic spectral activity is present at a position in the voice data where the pause should be, such as in the case of voice data, and if the dynamic spectral activity has an energy content that is

> within a preset energy range of the keyword energy content, then it is determined that no pause has occurred."

**Exh. 1**, '630 patent at Col. 10, *ll.* 43-52.

The PTO examiner agreed that, in the claimed method, a pause is detected by looking for changing spectral content, not by comparing the spectral content with some other value, such as background noise:

> Bellegarda *et al.* defines the process of a *pause* being detected as a silence between commands (column 3 lines 24-30), which would have made it obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to interpret *a lack of dynamic spectral activity* of a time as indicative of such a *silence*.

**Exh. 2**, Prosecution History at PH 69, ¶ 21 (italics in original).

Accordingly, "preventing operation of the electrical device when the spectral content is dynamic" can only mean what it says and does not add the concept of comparing the spectral content to some other value, such as the spectral content of background noise.

## III. THE TERM "ACTIVATING" MEANS TURNING ON

Claims 7 and 16 are directed to methods "for activating an electrical device..." The plain meaning of the term "activating" is to make active, or to turn on. *See* **Exh. 4**, sample of dictionary definitions. For example, the Cambridge Dictionary of American English defines "activating" as follows: "to cause (something) to start working." VST, however, contends that, within the field of speech recognition, "activating" also can mean turning off an appliance or making it less active. VST's *Markman* brief at 7. Contrary to VST's assertions, as Mr. Balentine explains, the term "activating" has no formal definition within the field of speech recognition. *See* Balentine II at ¶ 19. Moreover, had the inventors wished to claim these other functions, they could have used a broader term, such as "operating," Thus, activating cannot also mean turning

off an appliance. That is what the word "deactivating" means, and that term does not appear in the claims.

## IV. CLAIM 16 IS FATALLY INDEFINITE

A patent must conclude with "one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention." 35 U.S.C. § 112, ¶ 2. A claim that does not particularly and distinctly point out the invention is indefinite and thus invalid. Indefiniteness presents a question of law that is often resolved during "the court's performance of its duty as the construer of patent claims." *Personalized Media Communications, LLC v. Int'l Trade Comm'n*, 161 F.3d 696, 705 (Fed. Cir. 1998).

In *Group One, Ltd. V. Hallmark Cards, Inc.*, 2005 WL 1138998 (Fed. Cir., May 16, 2005), the Federal Circuit held that "A district court can correct a patent only if (1) the correction is not subject to reasonable debate based on consideration of the claim language and the specification and (2) the prosecution history does not suggest a different interpretation of the claims." *Id.* at *3 (citation omitted). Claim 16 cannot satisfy either prong of the required test.

There are two problems with Claim 16. First, the "receiving" step does not match the "recording" step. The claim requires the recording of "a command word portion and a pause portion." The receiving step, however, calls for "first and second command word portions" and three pause portions. The "comparing" step then requires comparing "said command word portion and said pause portion of said at least one received audio command" with the command and word portion of the recorded speech recognition data. It is not clear, however, which "said command word portion and said pause portion" received from the user should be compared to the speech recognition data because there are two word portions and three pauses. This

comparing step cannot be clarified simply by pluralizing "portion" because there is still only one recorded word and one pause portion in the speech recognition data.

Even if this discrepancy between the recording and receiving steps could be rectified, the claim does not satisfy the second prong of the *Group One* test because the prosecution history demonstrates that Claim 16 should have many more corrections made to it. The PTO examiner stated that original application claim 17 was allowable over the prior art "if rewritten in independent form including all of the limitations of the base claim and any intervening claims." **Exh. 2**, Prosecution History at PH 73, ¶ 34. In response, the inventors added new claim 27 (which ultimately issued as the present Claim 16) and stated that it represented "allowable claim 17 re-written in independent form." *Id.* at PH 90. But that is not true because, when the inventors added new application claim 27, they omitted several limitations from the original intervening claims 11, 13 and 14, such as "word entrance penalty," "background noise data," "command word score," and "background noise score." A comparison of those applications claims with Claim 16 (formerly application claim 27) proves this point. Claim 16 lacks these and other limitations from the original intervening claims and thus does not comply with the PTO examiner's directions. It is not immediately clear from the face of the patent how to correct this deficiency. Accordingly, this Court may not correct Claim 16, leaving it indefinite.

## V. VST SHOULD NOT BE ALLOWED TO RAISE OTHER CLAIMS

VST argues that it may assert additional patent claims against ScanSoft and, indeed, raises Claim 14 for the first time in its *Markman* brief. But VST did not expressly assert these other claims in answers to interrogatories, and this Court's order of April 13, 2005 clarifying the procedures for claim construction briefing precludes the assertion of such other claims.

Moreover, VST has argued the construction of Claim 14 in a brief that is too long and that should have thus been confined to a discussion of Claims 7 and 16 alone. Specifically VST's *Markman* brief is 26 pages, even though VST neither conferred with ScanSoft nor first sought leave of this Court before filing a brief that exceeds the 20-page limit of Local Rule 7.1(B)(4) (briefs "shall not, without leave of court, exceed twenty (20) pages, double-spaced"). *See Shabazz v. Cole,* 69 F. Supp. 2d 177, 186 (D. Mass. 1999) (Bowler, M.J.) (citing defendants' noncompliance with page limit requirements as an example of its pattern of disregard for the Local Rules). ScanSoft thus requests that, as a sanction, this Court ignore VST's discussion of Claim 14. That discussion takes up 5 pages. Thus, striking that discussion will leave the brief sufficiently close to the 20-page limit. In the alternative, ScanSoft respectfully requests leave to file a supplemental brief on any additional disputed terms of Claim 14.

## CONCLUSION

For the reasons stated above, ScanSoft respectfully requests that this Court adopt ScanSoft's proposed claim construction, which is firmly grounded in the instrinsic evidence.

Dated: June 3, 2005

SCANSOFT, INC.,

By its attorneys,

/s/ Erik P. Belt
Lee Carl Bromberg, BBO # 058480
Robert Asher, BBO # 022865
Erik Paul Belt, BBO # 558620
Lisa M. Fleming, BBO # 546148
Jack C. Schecter, BBO # 652349
BROMBERG & SUNSTEIN LLP
125 Summer Street
Boston, Massachusetts 02110-1618
Tel: (617) 443-9292
E-mail: ebelt@bromsun.com